Foundations and Applications in Large-scale AI Models: Pre-training, Fine-tuning, and Prompt-based Learning

Derek Cheng Google Research Mountain View, California, USA zcheng@google.com

Ed H. Chi Google Research Mountain View, California, USA edchi@google.com Linsey Pang Salesforce San Francisco, California, USA panglinsey@gmail.com

Wei Liu University of Technology Sydney Sydney, NSW, Australia weiliu.au@gmail.com Kexin Xie Salesforce San Francisco, California, USA

James Bailey University of Melbourne Victoria, Australia baileyj@unimelb.edu.au

kexin.xie@salesforce.com

ABSTRACT

Deep learning techniques have advanced rapidly in recent years, leading to significant progress in pre-trained and fine-tuned largescale AI models. For example, in the natural language processing domain, the traditional "pre-train, fine-tune" paradigm is shifting towards the "pre-train, prompt, and predict" paradigm, which has achieved great success on many tasks across different application domains such as ChatGPT/BARD for Conversational AI and P5 for a unified recommendation system. Moreover, there has been a growing interest in models that combine vision and language modalities (vision-language models) which are applied to tasks like Visual Captioning/Generation.

Considering the recent technological revolution, it is essential to have a workshop at the KDD conference that emphasizes these paradigm shifts and highlights the paradigms with the potential to solve different tasks. This workshop will provide a platform for academic and industrial researchers to showcase their latest work, share research ideas, discuss various challenges, and identify areas where further research is needed in pre-training, fine-tuning, and prompt-learning methods for large-scale AI models. The workshop will also foster the development of a strong research community focused on solving challenges related to large-scale AI models, providing superior and impactful strategies that can change people's lives in the future.

1 WORKSHOP ORGANIZERS

- Name: Derek Cheng; Affiliations: Google Research, Mountain View, California, USA; Contact: zcheng@google.com,
- Name: Linsey Pang; Affiliations: Salesforce, San Francisco, California, USA; Contact: panglinsey@gmail.com
- Name: Kexin Xie; Affiliations:Salesforce, San Francisco, California, USA; Contact: kexin.xie@salesforce.com
- Name: Ed H. Chi; Affiliations: Google Research, Mountain View, California, USA; Contact: edchi@google.com
- Name: Wei Liu (MAIN CONTACT); Affiliations: University of Technology Sydney, Sydney, NSW, Australia; Contact: weiliu.au@gmail.com

• Name: James Bailey; Affiliations: The University of Melbourne, VIC, Australia; Contact: <u>baileyj@unimelb.edu.au</u>

2 TOPIC

Pre-trained deep learning models, such as transformer-based language and vision models, have achieved tremendous success in many natural language processing (NLP) and computation vision (CV) tasks. They have been applied in various downstream application domains, for instance, question answering (QA), text and image generation, image classification, object detection, etc. Pre-trained models learn universal feature representations from large volumes of text/video/audio data using supervised learning or self-supervised learning and transfer this knowledge to downstream tasks, which avoids the training of downstream models from scratch.

For achieving the objectives of downstream tasks, the paradigm standard is shifting from "pre-train, fine-tune" to "pre-train, prompt, and predict" [4]. Due to this rapid growth of the modern pre-trained large models, there is a need to bring professionals together from both academic research and industry to present challenges in KDD community and discuss together to improve various current solutions, such as fine-tuning strategies, data sampling techniques, prompting methods, deployment scalability etc, which will create new challenges and new requirements for large pre-trained models in both theories and applications. This leads to the primary motivations of this workshop.

Some key challenges and new trends are listed in below:

• In the natural language processing (NLP) community, due to the power of generative features [1], large language models (LLM) is transforming and changing industries and individual lives. A recent survey found that 60% of tech leaders said that their budgets for AI language technologies increased by at least 10% in 2020 while 33% reported a 30% increase¹. In addition, the tech giant Microsoft, announced a new multiyear, multibillion-dollar investment with the artificial intelligence lab OpenAI, the creator of ChatGPT². Additionally, Google has recently also announced BARD: conversational

¹https://www.datanami.com/2023/01/10/large-language-models-in-2023-worth-the-hype/

²https://www.cnbc.com/2023/01/23/microsoft-announces-multibillion-dollarinvestment-in-chatgpt-maker-openai.html

AI services powered by the large language model LaMDA [7]. BARD competes directly with rival ChatGPT. Besides, Deepmind introduced Sparrow [3] in September 2022, which is considered similar to ChatGPT and BARD.

- In the computer vision (CV) community, the breakthroughs from deep Transformer models in NLP domain has attracted great interest to adapt them for vision and multi-modal learning tasks. The frequency of searching keywords, such as BERT, Self-Attention, and Transformers, appear in CV literature has shown consistent growth in recent years. For instance, the Vision Transformer (ViT) [2] marks the first step towards the merger of vision and language fields into a single unified model. It is the first time in the history of machine learning that a single model architecture has come to dominate both language and vision.
- In multi-modal settings, vision-and-language models like CLIP (Contrastive Language–Image Pre-training) [6] has demonstrated a strong zero-shot capability on various vision task and are used for many task-specific fine-tuning and downstream tasks. For instance, CLIPSeg [5] is a model that uses CLIP representations to create image segmentation masks. It achieved zero-shot image segmentation by training a Transformer-based decoder on top of the CLIP model. The pretrain-then-transfer learning approach to Vision-language tasks has become the de facto standard due to the solid representational power of these models trained on large-scale data.

Although we have seen all of the great progress of large pretrained language and vision models, many open challenges and limitations of these models are yet to be tackled and addressed, for instance:

- As of now, there are many top-leading language models in the NLP community, what are the connections and differences across them? Is there any fundamental architectural change since the publication of 2017's "Attention is all you need" [8]? If the higher performance is mostly due to the large scale of training data and the domain specificity of the data, then how to minimize training and fine-tuning costs across domains?
- Although prompt-based learning [4] becomes a popular strategy for allowing pre-trained AI models to be re-purposed for downstream tasks without additional training, is promptbased learning better than fine-tuning in every aspect? What are the pros and cons of prompt learning?
- How to reduce the built-in bias in pre-trained models and prevent the harm generated from the models in the real world, for instance, what if ChatGPT generates wrong health advice to patients?
- How would large language models revolutionize traditional search and recommendation systems? Modern search and recommendation systems have moved towards a more complex system, for instance, data-efficient Prompt-based NLP (Pre-train, Prompt, Predict) recommender systems (P5) are to unify recommendation tasks with one model, one loss, and one data format. With these advancements powered by

Derek Cheng, Linsey Pang, Kexin Xie, Ed H. Chi, Wei Liu, and James Bailey

large-scale models, new challenges and vital new requirements for search and recommendation systems will emerge to be solved.

3 WORKSHOP RELEVANCE

KDD is a top conference that provides the most premier international forum for presentations and discussions of research advances in data mining and machine learning.

Given the rapid development of text/image/video techniques, there has been significant progress on the pre-trained and finetuned models. For instance, in the NLP domain, since 2019, the language model has shifted from a fully supervised paradigm to a pre-trained and fine-tuned paradigm. Currently, the "pre-train, fine-tune" procedure is shifting to the "pre-train, prompt, and predict" paradigm [4]. Due to this substantial change, it is necessary to discuss the recent advances in pre-trained models, through a well-organized workshop at a top-tier conference like KDD. KDD provides a platform for researchers and practitioners to share stateof-art solutions, and to discuss future directions and challenges from data mining and machine learning view. Therefore, our workshop, titled "Pre-train and Fine-tune Large-scale Models: theoretical solutions, applications and challenges", is highly relevant to the KDD conference.

4 TARGET AUDIENCE

Currently, the dramatic growth of the pre-trained and fine-tuning large CV and NLP models poses great interest to the research and industry community, which calls for novel and improved solutions to solve open challenges and create new opportunities, for instance: how to improve the performance of pre-training models and finetuning models on poorly curated data, and in what configurations should we expect pre-training to be more effective than learning from scratch, etc.

This workshop will bring academic and industrial practitioners together who have strong interests in studying various techniques to pre-trained and/or fine-tuned large-scale CV and NLP models to discuss and learn the latest trends. This includes both theoretical and application challenges. We are expecting at least 30-50 people to join this workshop.

5 MOTIVATION, GOAL AND EXPECTED OUTCOMES

Although pre-trained and fine-tuned large-scale models have progressed significantly over recent years, increasingly more challenges are still needed to be addressed. For instance, one of the key challenges is how to better use the knowledge in pre-training from prompt learning to building foundation models. This workshop will provide academic and industrial researchers with a platform to present their latest work, share research ideas, present and discuss various challenges, and identify the areas where further research is needed for improving larger-scale pre-train, fine-tune strategies. It will foster the development of a strong research community focused on solving pre-training and/or fine-tuning models at large-scale.

This workshop is expected to have submissions from researchers in both academia and industry broadly from data science, NLP, computer vision, and other areas related to pre-trained and finetuned large-scale models.

6 WORKSHOP THEMES AND TOPICS

The Workshop will serve as a platform for publication and discussion of the theoretical and application challenges and approaches of pre-trained, fine-tuned models, etc.

We invite submissions of long (eight papers) and short (four pages) papers, representing original research, preliminary research results, and proposals for new work in academia or industry. All submissions will be single-blind and will be peer-reviewed by an international program committee of researchers and industrial professionals and experts.

Accepted submissions will be required to be presented at the workshop and will be published in a dedicated workshop proceeding by the workshop organisers.

Topics of interest in this workshop include but are not limited to:

- Pre-training
 - Improvements in pre-training: supervised pre-training, self-supervised pre-training with various auxiliary tasks, meta-learning, prompt-based Learning, multi-modal pretraining etc.
 - Novel pre-training methods to maximize generalization
 - Model selection for pre-trained models
 - Pre-training for various application domains, such as computer vision, natural language processing, robotics, etc
- Fine-tuning:
 - Domain/task adaptive fine-tuning
 - Intermediate-task, multi-task, self-supervised, MLM finetuning
 - Parameter-efficient fine-tuning: sparse parameter tuning, pruning
 - Text-to-Text, Text-to-image, Image-to-text, multi-modal fine-tuning, effectively using large autoregressive pretrained models
 - Fine-tuning for various application domains, such as computer vision, natural language processing, robotics, etc
- Prompted/Instruction-based:
 - Manual Template Engineering
 - Automated Template Learning
 - Multi-Prompt Learning; Multi-tasks instruction tuning
 - Instruction tuning with HF/RLHF
 - chain-of-thought (CoT) prompting
- Performance:
 - Model compression techniques
 - Large-scale model deployments
 - Efficient and effective training/inference
 - Empirical analysis of various pre-training and fine-tuning methods
 - Generalization bounds of different pre-training and finetuning methods
 - Stability, sparsity and robustness strategies
- Downstream tasks of large-scale models:
 - NLP models for Text Generation
 - NLP models for Text Summarization

- NLP models for Question Answering
- NLP models for other downstream tasks
- CV models for Image Captioning
- CV models for Semantic Segmentation
- CV models for Object Tracking
- CV models for other downstream tasks
- Applications powered by large-scale models:
 - Conversational AI, Conversational Chatbots
 - Enhanced Web Search, Search Engine
 - Unified, Personalized next generation recommender systems

7 PROGRAM SKETCH (HALF-DAY ~3 HOURS)

The proposed workshop will have the following events to fulfill the purpose:

- **Invited talks (1.5 hours)**: We plan to invite 3-4 speakers (20-30 mins each) from both industry and academia to give informative presentations related to our focused topics.
- **Paper presentations (1 hour)**: 4 selected long papers will be invited to give a 15 min talk each.
- **Poster session (30 mins)**: We will host a poster session including both accepted long and short papers. The poster session will help participants to have a deeper discussion on various works and gain more exposure to their work.

Agenda

- Opening Remarks: 5 mins
- Keynote 1: 30 mins (talk + Q&A)
- Keynote 2: 30 mins (talk + \widetilde{Q} &A)
- Contribution talk 1: 15 min
- Contribution talk 2: 15 min
- Break: 10 mins
- Invited talk: 20 mins (talk + O&A)
- Invited talk: 20 mins (talk + Q&A)
- Contribution talk 3: 15 min
- Contribution talk 4: 15 min
- Poster Session: 30 min

Tentative Invited Speakers :

- Yongfeng Zhang, Rutgers University
- Sanjay Chawla, Qatar Computing Research Institute
- Niels Rogge, Huggingface
- Yu Zheng, JD.COM

Tentative list of Program committee:

- Yuxuan Liang, National University of Singapore
- Yanjie Fu University of Central Florida
- James Allan, University of Massachusetts Amherst
- Evangelos Kanoulas, University of Amsterdam
- Zhensong Qian, Amazon
- Vinayak Gupta, IBM
- Tesi Xiao, University of California, Davis
- Philippe Laban, Salesforce
- Xue Li, Microsoft

Encourage physical attendance and in-person activities: The organizers will encourage physical attendance and in-person activities in the workshop via the following:

- Invite high-profile and famous senior academics or industry leaders to be keynote speakers
- Require all accepted workshop papers to be presented in the workshop in person
- Set up a Best Paper Award in the workshop and announce the winner in the workshop.

8 PREVIOUS ITERATION OF THE WORKSHOP

Relevant workshops run by some of us were held at KDD 2021, KDD 2022, WSDM 2022, and CIKM 2022.

- Deep Learning for Search Recommendation workshop is a full-day workshop in conjunction with CIKM 2022. The workshop (https://dl4sr.github.io/dl4sr22/) is organized by: Wei Liu (University of Technology Sydney), Kexin Xie (Salesforce), Linsey Pang (Salesforce), James Bailey (The University of Melbourne), Longbing Cao (University of Technology Sydney), Yuxi Zhang (Salesforce). The workshop invited 5 keynote speakers: Ed H. Chi (Google), Caiming Xiong (Salesforce), Xiuzhen (Jenny) Zhang (RMIT University), Belinda Zeng (Amazon), Jonathan Purnell (Spectrum Labs). There is a total of 27 submissions, among which 8 long papers and 4 short papers were accepted.
- The 15th AdKDD workshop (https://www.adkdd.org/) is a full-day workshop in conjunction with KDD 2022. The workshop is organized by: Abraham Bagherjeiran (eBay), Kuangchih Lee (Alibaba), Kun Liu (Amazon), Wei Liu (University of Technology Sydney), Kexin Xie (Salesforce) , Linsey Pang (Salesforce), Mihajlo Grbovic (AirBnB), Nemanja Djuric (Aurora Innovation), Suju Rajan (Amazon), Vladan Radosavljevic (Spotify). The AdKDD workshops held in conjunction with KDD conferences in the past years (2017-2021) continue to generate interest from academia and industry, as one of the top venues specifically for advertising research.
- 1st International Workshop on Interactive and scalable information retrieval methods for e-commerce (https://isirecom.github.io/) was held in conjunction with WSDM 2022. Around 30-35 attendees joined this half-day workshop. The workshop invited 3 keynote speakers (Maarten de Rijke from University of Amsterdam, Khalifeh AlJadda from HomeDepot, Xiaokui Xiao from National University of Singapore and Yanjie Fu from University of Central Florida). The workshop was co-organized by: George Karypis (University of Minnesota Twin Cities), Hamed Zamani (University of Massachusetts Amherst), Lingfei Wu (JD.COM Silicon Valley Research Center), Xiquan Cui (The HomeDepot), Linsey Pang (Walmart Labs), Vachik Dave (Walmart Labs). There were a total 15 paper submissions, among which 3 long papers and 2 short papers were accepted for oral presentations. Submissions were from both academia and industry, for instance: purdue university, amazon search etc. The difference between this workshop and the proposed one is that, this workshop focuses mainly on information retrieval methods, while the proposed workshop focuses on web data mining using deep learning-based methods.
- 2nd International Workshop on Industrial Recommendation Systems (https://irsworkshop.github.io/2021/) was held as

a full-day workshop in conjunction with KDD 2021. There were 23 paper submissions, among which 5 papers were selected for oral presentations, and 6 papers were selected for posters. Submissions were from both academia and industry, including Yale University, UCLA, Amazon, eBay, Walmart, etc. The workshop invited 3 keynote speakers (Xiangnan He from UMN/Amazon, Christoph Koflerg from Netflix, Rishabh Mehrotra from Spotify, Anoop Deoras from Amazon, Even Oldridge from Nvidia and Daxu from Walmartlabs). The workshop was also supported by our co-organizers, including Philip S. Yu (University of Illinois at Chicago), George Karypis (University of Minnesota Twin Cities), Dawei Yin (Baidu), Justin Basilico (Netflix), Mohit Sharma (Youtube) , Lingfei Wu (JD.COM Silicon Valley Research Center), Linsey Pang (WalmartLabs) and Jianpeng Xu (WalmartLabs).

9 POSSIBLE MARKETING VENUES

We are planning to publicize the workshop using:

- Social media and Apps (LinkedIn, Twitter, Facebook, WeChat etc.)
- Targeted research community using email lists such as MLnews, DBWorld, WikiCFP, etc.
- Spreading the word between industrial and academia connections.

10 ORGANIZER PROFILES

- **Dr. Derek Cheng** is a senior staff software engineer and tech lead manager at Google Research, where he builds works on representation learning research for recommenders and its applications in Google Ads, Play, and Google News. Prior to joining Google in 2013, he finished his Ph.D. in Texas AM University with a focus on geo-social data mining. Derek has over 20 peer-reviewed articles published in prestigious conferences and journals for information retrieval and data mining. He won the test of time award at CIKM, and co-chaired the industry day in WSDM 2020.
- Dr. Linsey Pang is a Principal Applied Scientist, Salesforce. She is working on text and image related deep learning RD modeling projects. Prior to this, she was Principal Data Scientist at Walmart Lab. She was leading the Pricing, One demand data science projects in the Merchant Technology organization. Prior to joining Walmart Lab, she was working as an applied scientist at eBay Inc. She is a co-organizer of DL4SR 2022, AMLTS 2022 workshops collocated with CIKM 2022, AdKDD 2022 workshop collocated with KDD 2022, Isir-ecom 2022 workshop collocated with WSDM 2022, IRS 2021 workshop collocated with KDD 2021.
- Dr. Kexin Xie is a Vice President of Engineering at Salesforce, responsible for data science research, ML engineering and architecture. He leads the team to push forward the AI initiative at very large scale, the data system making 160 billion predictions on a daily basis. Kexin has published many papers in top-tier computer science journals and conferences such as ACM Conference on Recommender Systems, ACM Transactions on Database Systems (TODS), Very Large Database Journal (VLDB), and was invited reviewer for

many prestigious computer science publications such as IEEE Transactions on Knowledge and Data Engineering (TKDE), the VLDB Journal, and Journal of Information Sciences. He is speaking regularly at tech conferences like Spark Summit '17, '18. AI Next Con '19, Strange Loop '19, and ODSC '19, WSDM '22. He co-organized workshops conjunction with KDD '22 and CIKM '22.

- Dr. Ed H. Chi is a Distinguished Scientist at Google, leading several machine learning research teams focusing on neural modeling, reinforcement learning, dialog chatbot models called LaMDA, reliable/robust machine learning, and recommendation systems in Google Brain team. His team has delivered significant improvements for YouTube, News, Ads, Google Play Store at Google with 600 product improvements since 2013. With 39 patents and >150 research articles, he is also known for research on user behavior in web and social media. Prior to Google, he was the Area Manager and a Principal Scientist at Palo Alto Research Center's Augmented Social Cognition Group, where he led the team in understanding how social systems help groups of people to remember, think and reason. Ed completed his three degrees (B.S., M.S., and Ph.D.) in 6.5 years from University of Minnesota. Recognized as an ACM Distinguished Scientist and elected into the CHI Academy, he recently received a 20-year Test of Time award for research in information visualization. He has been featured and quoted in the press, including the Economist, Time Magazine, LA Times, and the Associated Press. An avid swimmer, photographer and snowboarder in his spare time, he also has a blackbelt in Taekwondo.
- A/Prof. Wei Liu is an Associate Professor in Machine Learning, and the Director of Future Intelligence Research Lab, in the School of Computer Science, the University of Technology Sydney (UTS). He obtained his PhD degree in Machine Learning research at the University of Sydney (USyd). His current research focuses are adversarial machine learning, cybersecurity, game theory, multimodal machine learning, natural language processing, and intrusion detection. Wei has published over 100 papers in CORE A*/A and Q1 (i.e., top-prestigious) journals and conferences. He has received 3 Best Paper Awards. In addition, one of his first-authored papers received the Most Influential Paper Award in conference PAKDD 2021. Wei served as a tutorial chair at ICDM 2021. He is a co-organizer of the NeuroRec 2021 workshop collocated with ICDM 2021, the AdKDD 2022 workshop collocated with KDD 2022, and the DL4SR 2022 workshop collocated with CIKM 2022.
- **Prof. James Bailey** is currently a Professor with the Melbourne School of Engineering, The University of Melbourne, and the Program Lead of Artificial Intelligence. He was an Australian Research Council Future Fellow. He is a Researcher in the field of machine learning and artificial intelligence, including interdisciplinary applications and operational frameworks. His interests particularly relate to the assurance, certification, and safety of systems based on machine learning and artificial intelligence.

James serves in the editorial boards of Journal of Artificial Intelligence Research (JAIR) and the ACM Transactions on Intelligent Systems and Technology (ACM TIST). He has previously served on the editorial board of the IEEE Transaction on Data Engineering (IEEE TKDE). He was a co-PC Chair for the 20th IEEE International Conference on Data Mining (ICDM 2021); a co-PC Chair for the 32nd Australasian Joint Conference on Artificial Intelligence (AI 2019); a co-PC Chair for the 20th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2016); and a co-General Chair for the 24th ACM International Conference on Information and Knowledge Management (CIKM 2015).

REFERENCES

- [1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. Advances in neural information processing systems 33 (2020), 1877–1901.
- [2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2010. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv 2020. arXiv preprint arXiv:2010.11929 (2010).
- [3] Amelia Glaese, Nat McAleese, Maja Trębacz, John Aslanides, Vlad Firoiu, Timo Ewalds, Maribeth Rauh, Laura Weidinger, Martin Chadwick, Phoebe Thacker, et al. 2022. Improving alignment of dialogue agents via targeted human judgements. arXiv preprint arXiv:2209.14375 (2022).
- [4] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.
- [5] Timo Lüddecke and Alexander Ecker. 2022. Image segmentation using text and image prompts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 7086–7096.
- [6] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.
- [7] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, et al. 2022. Lamda: Language models for dialog applications. arXiv preprint arXiv:2201.08239 (2022).
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. Advances in neural information processing systems 30 (2017).